On media coverage of the COVID-19 outbreak: A corpus-based collocation study

Yejin Kim
(Seoul National University)

Kim, Yejin. 2020. On media coverage of the COVID-19 outbreak: A corpus-based collocation study. *SNU Working Papers in English Linguistics and Language 17, 47-73.* This preliminary study explores the diverse ways in which the ongoing COVID-19 outbreak has impacted every sphere of our lives in terms of the ways it is conceptualized in the contemporary media. By way of a segmented collocation study using a recently launched Coronavirus Corpus, this study pulls into view the particularized characterization of the novel Coronavirus vis-à-vis traditional virus description. The results rendered from the present study allude to the way how our general perception about the COVID-19 outbreak could substantially be informed by and fed back into such media-circulated language usage. Ultimately, it hopes to accentuate the importance of cultivating critical awareness towards the consumption of media news in times of unprecedented crisis. **(Seoul National University)**

Keywords: COVID-19 outbreak, Coronavirus, media discourse analysis, sociolinguistics, corpus collocation study

1. Introduction

1.1. Motivation

The motivation behind this preliminary research arose out of practical concerns, namely the need to address the implications of the COVID-19 outbreak which is massively and indefinitely impacting our everyday lives. Since the way we perceive disease outbreaks to which we lack primary access is predominantly shaped by media coverage, the global Coronavirus epidemic is unprecedented in that it is still ongoing and that it has far reaching repercussions to the extent that it is often called an incident that had already catapulted us into a new post-corona era. It is thus with this aim in mind that I set out to carry out a corpus-based collocation study on the media coverage of the Coronavirus, which I

believe warrants more empirical coverage that could potentially open up some other newsworthy avenues of investigation as regards general disease media coverage. Additionally, considering the fact that the media is the primary source from which the Coronavirus *discourse* is propagated to inform our perception, its leading role in the characterization of the current phenomenon is an area of inquiry that commands our better attention. Since the media is naturally disposed to present either a magnified or a reductive view of any social phenomena, cultivating critical awareness becomes an imperative if we are to maintain a more balanced view instead of indiscriminately consuming skewed ones.

1.2. Recent Documentation on the Coronavirus Outbreak

Since the Coronavirus outbreak is a fairly recent incident but a one with infinitely devastating ramifications, scholars and professionals in diverse fields have rapidly undertaken to investigate its social impact, mostly in terms of the dissemination and consumption of misinformation circulating around the social media. It is reported that there is a strong correlation between right-leaning media consumption and misinformation endorsement (Motta et al., 2020). A host of other studies have been carried out in a similar vein, which collectively raised alarm about the spread of unverifiable information on social media platform that could seriously endanger public health (Kouzy et al., 2020; Binjola & Patel, 2020). Most of them pointed to Twitter as the biggest locus of production and reproduction of such misinformation (Sharma et al., 2020) or a ‘hoaxes’ in the words of Rahardi (2020). Some have broached the surge of misinformation from a risk communication prospective and argued that this misinformation is a superordinate meta-risk that compounds the already existing health risk posed by the Coronavirus outbreak (Krause et al., 2020). Others have refuted it by claiming that a considerable amount of evidence-based

information was relatively longer-lasting (i.e. retweeted more) in drawing engagement on the social media (Pulido et al., 2020). In an extended study that investigated various online advertisement domains incorporating the Coronavirus, the possibility of another deluge of misinformation has been raised significantly, which serves as a clarion call for the media as regards their principal role in quality control for constructive public health discourse (Mejova & Kalimeril, 2020). Another short commentary accentuated the role of medical librarians and media professionals as main distributors of information, who should also be the ones providing guidance through teaching information literacy in preparation for another similar crisis (Ashrafi-rizi & Kazempour, 2020). Last, Wen et al. (2020) argued that biased media coverage spawns unwarranted racial discrimination and stressed the importance of informed consumption of information generated by the media. All the recent documentations on the Coronavirus outbreak predominantly center around public health concerns and the circulation of misinformation on social media without any substantial attention given to the actual reporting of the Coronavirus, which precisely provides for my research to be a departing point for addressing the relationship between media discourse and disease perception.

2. Methodology

The data for this study have been collected from two major corpora under the BYU-BNC corpora accessible online. The first corpus used as a reference corpus is the NOW (News on the Web) corpus which contains 10.3 billion words of data from web-based newspapers and magazines from 2010 to the present time (the most recent update was checked to be 2020-6-24). More importantly, the corpus grows by about 180-200 million words of data each month (from about 300,000 new articles), or about two billion words each year, whose extensiveness of

data storage accordingly lends credibility to the generalizability of any consistently emerging pattern.

The key corpus is the recently launched Coronavirus Corpus which is designed to be the definite record of the social, cultural, and economic impact of the Coronavirus (COVID-19) in 2020 and beyond. Unlike resources like Google Trends (which just show what people are searching for), the corpus shows what people are actually saying in online newspapers and magazines in 20 different English-speaking countries. The corpus (first released in May 2020) is currently about 389 million words in size, and it continues to grow by 3-4 million words each day.

The main search feature used for the present study was the LIST search feature that provides the search syntax with the angled bracket [ ] and capitalization. This search feature facilitated the lemma search and allowed me to group together all the relevant inflected forms and organize this search into three subsections consisting of different part of speech collocations. Some other features such as the COLLOCATION were initially used to get a broader picture and then determine the most appropriate search strings for each part of speech. Although about a hundred collocations were retrieved for each section, only 30 tokens were treated as relevant and cited in this paper in view of their high frequencies and particularly, for space reasons.

3. Result & Discussion

Prior to proceeding with this section covering the results and discussion of the study, it should be noted that all the tables show the relative collocation rank between the key word (*the Coronavirus*) and the reference word (*the virus)*. All the charts show the distribution of the commonly occurring collocations identified from the previous tables and the figures plotted on the chart are presented in normalized

percentages for comparison. These normalized figures were produced by calculating the ratio of common collocation against the total collocation tokens of each part of speech retrieved from two corpora. Finally, all the lists simply reclassify the collocations that occur 'exclusively' in each corpus (highlighted in either green or red) to allow for subsequent qualitative analysis.

3.1. Noun Collocation

In order to observe collocation patterns within the particular noun part of speech domain, I searched two key noun phrase constructions that would allow me to observe any different collocations of our keyword *the Coronavirus* in relation to its reference word *the virus.* First, the postposed nouns were posited to be illustrating the usage of the keyword *the Coronavirus* vis-a-vis its reference word and the search window was extended to 2 so as to accommodate the next occurring noun as a complete compound noun. Initially, the search string '***the Coronavirus/virus*** **[NOUN]**' was entered into the LIST search feature, which yielded the immediately following nouns of each word. Next, to complement underspecification, another string **'*the Coronavirus/virus*** **[NOUN] [NOUN]'** has been entered using the same search feature, which will be equally dealt with in the following section. Last, another relevant noun phrase construction composed of '**[NOUN] [PREP]** ***the Coronavirus/virus*****'** was employed as a search string to identify any other missed noun collocations.

As you can see in the table 1 above, the collocations ranked high (ranked 1-10) constitute the majority of the tokens and are shown to share most nouns only with [shutdown] and [pandemic] occurring exclusively with *the Coronavirus* and [transmission] and [particle] occurring exclusively with *the virus.* As regards the quantitative pattern, 18 nouns out of a total of 30 nouns immediately followed both *the Coronavirus* and *the virus* but with varying degrees as made

Table 1. List of all collocations (sorted by frequency)

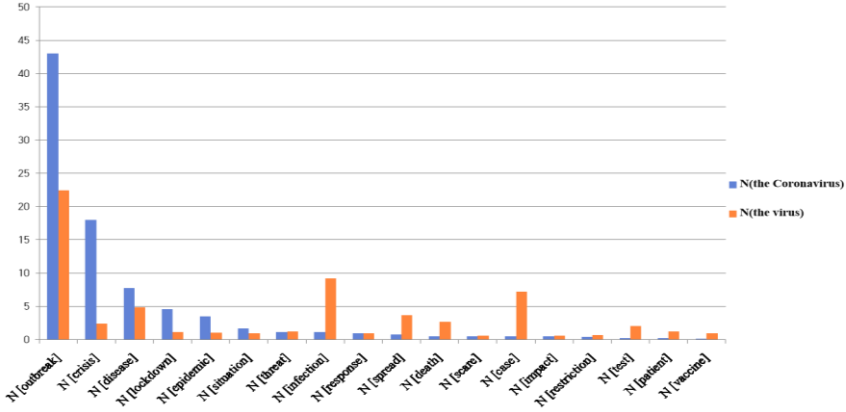
	the Coronavirus (Coronavirus corpus)		*the virus* (NOW corpus)	
	Types	**Tokens**	**Types**	**Tokens**
1	the Coronavirus [outbreak]	56879	the virus [outbreak]	12838
2	the Coronavirus [crisis]	24397	the virus [infection]	5310
3	the Coronavirus [disease]	10081	the virus [case]	4182
4	the Coronavirus [lockdown]	6119	the virus [disease]	2771
5	the Coronavirus [epidemic]	4577	the virus [spread]	2175
6	the Coronavirus [situation]	2291	the virus [transmission]	1962
7	the Coronavirus [shutdown]	2179	the virus [death]	1571
8	the Coronavirus [threat]	1552	the virus [crisis]	1409
9	the Coronavirus [infection]	1456	the virus [particle]	1242
10	the Coronavirus [pandemic]	1210	the virus [test]	1201
11	the Coronavirus [response]	1206	the virus [strain]	871
12	the Coronavirus [spread]	1146	the virus [threat]	721
13	the Coronavirus [emergency]	955	the virus [patient]	710
14	the Coronavirus [death]	781	the virus [lockdown]	643
15	the Coronavirus [scare]	733	the virus [epidemic]	605
16	the Coronavirus [case]	727	the virus [response]	572
17	the Coronavirus [impact]	661	the virus [situation]	554
18	the Coronavirus [restriction]	580	the virus [vaccine]	546
19	the Coronavirus [curve]	341	the virus [research]	482
20	the Coronavirus [test]	327	the virus [type]	451
21	the Coronavirus [patient]	314	the virus [concern]	434
22	the Coronavirus [era]	312	the virus [restriction]	432
23	the Coronavirus [quarantine]	285	the virus [attack]	422
24	the Ccoronavirus [fight]	259	the virus [scare]	382
25	the Ccoronavirus [issue]	209	the virus [symptom]	354
26	the Coronavirus [vaccine]	204	the virus [impact]	351
27	the Coronavirus [recession]	172	the virus [protection]	350
28	the Coronavirus [fallout]	170	the virus [containment]	309
29	the Coronavirus [panic]	164	the virus [replication]	302
30	the Coronavirus [downturn]	157	the virus [sample]	291
Total		130743		57303

※ word strings (yellow): commonly identified collocations
word strings (green): collocations exclusive to the keyword *'the Coronavirus'*
words strings (red) : collocations exclusive to the reference word *'the virus'*

apparent by different orderings among them. The remaining 12 nouns were sorted as occurring exclusively with either *the Coronavirus* or *the virus*, which are highlighted in green (left) and red (right) respectively to aid comparison.

Additionally, the noticeably different orderings even within the commonly identified collocations (in yellow) were scrutinized further. These common collocations are at first sight, least common denominators that substantiate the inherent features of the virus. By deriving the percentage of these common collocations from normalizing each corpus, a significant difference even within these common collocations is observed, as demonstrated in the chart below.

Chart 1. Commonly identified collocations (words in yellow, sorted by ratio)



Based on this chart 1, it appears that both *the Coronavirus* and *the virus* share [outbreak] as their closest collocate but with conspicuously different ratios. Also, *the Coronavirus* shares relatively stronger semantic affinity with the nouns [crisis], [disease], [lockdown] and [epidemic] while its reference word *the virus* shares closer semantic affinity with remaining nouns such as [infection], [spread], [death], [cases], [test], [patient] and [vaccine], all of which are typical words employed in the 'clinical' description of the virus itself. This different collocation pattern between the two words is more sharply delineated if we look at the list below that shows the collocations exclusive to each word.

List 1. Exclusive collocations (words in green & red)

the Coronavirus	shutdown, pandemic, emergency, curve, era, quarantine, fight, issue, recession, fallout, panic, downturn
the virus	transmission, particle, strain, research, type, concern, attack, symptom, protection, containment, replication, sample

The list 1 above makes salient the difference valence that each word exhibits. When we look at the nouns that exclusively collocate with *the Coronavirus*, they are nouns that generally refer to the social

repercussions of the Coronavirus such as [shutdown], [pandemic], [emergency], [era], [recession], [fallout], [panic] and [downturn]. Note how some words such as [era], [recession], or [fallout] can also be easily modified by other adjectives such as 'post-modern', 'economic', or 'radioactive/environmental' that could alternatively take up the same slot that the keyword *the Coronavirus* occupies. Therefore, the replaceable word *Coronavirus* in these types of noun collocations has a weaker semantic contribution in relation to its immediately following nouns in that it is actually the immediately following nouns that retroactively characterize the Coronavirus as a vector of social change. Therefore, the implicature of 'causality' is insidiously wrought into these types of new noun collocations with the bridging word (i.e., Coronavirus-'*induced*') veiled in exchange. This is thrown in sharp contrast if we look at the relatively limited set of nouns that the reference word *the virus* collocates with, which demonstrates how *the virus* is a major semantic contributor in the formation of new noun collocations. The way the word *virus* effectively 'constrains' its collocating pattern is exemplified in the immediately following nouns such as [transmission], [strain], [research], [symptom], [replication], and [sample]. To sum up, the keyword *the Coronavirus* in these noun collocations only serves as a proxy for other candidate adjectives, which is borne out by its loose association with various nouns that do not necessarily relate to the features of the virus itself. It is therefore, a placeholder noun that indexes the 'causality' factor that eventually led to the emergence of new noun collocations.

3.1.2. *the Coronavirus/virus* + [COMPOUND NOUN]

In this section, the search window has been extended to 2 in order to accommodate 'compound nouns' collocating with the two words. For example, [*the Coronavirus* lockdown] result token would yield all the tokens containing the word *lockdown* as their immediate collocate, with

the second following noun varying to a great extent such as in [lockdown *measure, relief, restrictions, period*] and etc. Also, a search window of 1 would not be conducive to characterizing the key and the reference word such as [*the Coronavirus* job], [*the Coronavirus* health], [*the Coronavirus* business], [*the Coronavirus* market], [*the Coronavirus* travel] and etc. Their meanings remain opaque if we leave out the crucial element constituting the compound noun (i.e., the next occurring noun), so the incorporation of these compound nouns was also a necessary procedure.

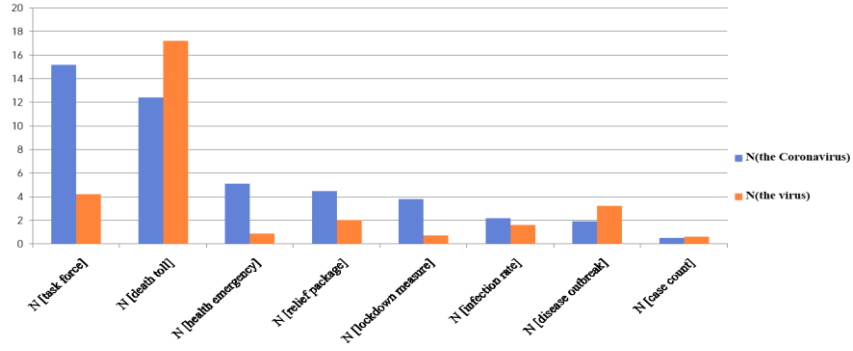
Table 2. List of all collocations (sorted by frequency)

	the Coronavirus (Coronavirus corpus)		*the virus* (NOW corpus)	
	Types	Tokens	Types	Tokens
1	the Coronavirus [task force]	834	the virus [death toll]	477
2	the Coronavirus [death toll]	683	the virus [testing kit]	173
3	the Coronavirus [job retention]	365	the virus [reference laboratory]	137
4	the Coronavirus [health emergency]	281	the virus [research institute]	124
5	the Coronavirus [health information]	266	the virus [task force]	119
6	the Coronavirus [relief package]	249	the virus [containment measure]	107
7	the Coronavirus [lockdown measure]	212	the virus [disease outbreak]	90
8	the Coronavirus [business interruption]	148	the virus [health authorities]	72
9	the Coronavirus [infection rate]	124	the virus [prevention measure]	71
10	the Coronavirus [public health]	118	the virus [relief package]	57
11	the Coronavirus [disease outbreak]	105	the virus [aid package]	52
12	the Coronavirus [stimulus package]	95	the virus [containment effort]	46
13	the Coronavirus [update newsletter]	81	the virus [infection rate]	45
14	the Coronavirus [outbreak impact]	47	the virus [vaccine market]	35
15	the Coronavirus [stay-at-home order]	45	the virus [disease case]	33
16	the Coronavirus [market crash]	43	the virus [reproduction rate]	31
17	the Coronavirus [travel ban]	40	the virus [news coverage]	29
18	the Coronavirus [shelter-in-place order]	37	the virus [life cycle]	29
19	the Coronavirus [alleviation program	33	the virus [health emergency]	27
20	the Coronavirus [emergency relief]	33	the virus [outbreak spread]	27
21	the Coronavirus [response coordinator]	32	the virus [incubation period]	25
22	the Coronavirus [food assistance]	31	the virus [filtration market]	23
23	the Coronavirus [stock market]	31	the virus [lockdown measure]	22
24	the Coronavirus [case count]	29	the virus [time bomb]	22
25	the Coronavirus [rescue package]	28	the virus [mitigation effort]	21
26	the Coronavirus [testing site]	27	the virus [prevention effort]	19
27	the Coronavirus [front line]	23	the virus [case count]	18
28	the Coronavirus [quarantine center]	23	the virus [control effort]	18
29	the Coronavirus [closure period]	22	the virus [vaccine candidate]	17
30	the Coronavirus [emergency response]	22	the virus [confinement measure]	17
Total		5474		2771

Overall, there were 8 compound noun collocations found to be commonly shared between the two words with 22 distinctive collocations identified. As you can see in the table 2 above, the top ten collocations that make up the majority of the tokens already differ in distribution. Only [task force], [death toll] and [relief package] occur as definitely common collocations while other commonly identified collocations such as [health emergency], [lockdown measure] and [infection rate] rank high only in the Coronavirus corpus. Therefore, the

bulk of the common collocations already occur predominantly in the Coronavirus corpus whereas the same common collocations are relatively sparsely and evenly distributed all throughout the NOW corpus. This observation suggests that there is already a noticeable valence difference between the keyword *the Coronavirus* and the reference word *the virus,* which is plotted in the chart below.

Chart 2. Commonly identified collocations (words in yellow, sorted by ratio)



If we take a closer look at this chart 2, the keyword *the Coronavirus* exhibits tighter semantic affinity with words such as [task force], [health emergency], [relief package] and [lockdown measure] while the reference word *virus* remains closely associated with words such as [death toll] and [disease outbreak] in comparison. This different collocation pattern is once again borne out if we look at the list below that distinguishes the compound nouns that occur exclusively with each word.

The contrasting compound noun collocation pattern from the list 2 above is already evocative of the findings from the previous section. First note how *the Coronavirus* easily collocates with other compound nouns such as [business interruption], [stimulus package], [market crash], [travel ban], [alleviation program], [emergency relief], [food

List 2. Exclusive collocations (words in green & red)

the Coronavirus	job retention, health information, business interruption, public health, stimulus package, update newsletter, outbreak impact, stay-at-home order, market crash, travel ban, shelter-in-place order, alleviation program, emergency relief, response coordinator, food assistance, stock market, rescue package, testing site, front line, quarantine center, closure period, emergency response
the virus	testing kit, reference laboratory, research institute, containment measure, health authorities, prevention measure, aid package, containment effort, vaccine market, disease case, reproduction rate, news coverage, life cycle, outbreak spread, incubation period, filtration market, time bomb, mitigation effort, prevention effort, control effort, vaccine candidate, confinement measure

assistance], [stock market] and [rescue package]. These types of compound nouns are the ones routinely cited in times of economic recessions and are in no need of further modification by *the Coronavirus*. Therefore, *the Coronavirus* as a noun acts as a stand-in to accentuate a certain property of the following noun, namely, its trigger. In this respect, its semantic contribution to the formation of new noun collocations is only nominal as it tends to be used as not so much a noun but as an adjective. This 'adjectivized' usage of the key word *the Coronavirus* becomes clear if we look at the other compound nouns that collocate exclusively with the reference word *the virus* (e.g., [research institute], [vaccine market], [reproduction rate], [life cycle], [incubation period]), whose occurrences are for the most part constrained by the features that the *virus* itself projects.

3.1.3. [NOUN] + [PREP] + *the Coronavirus/virus*

Overall, the results from searching this noun phrase construction revealed that there are 19 commonly shared collocation and 11 exclusive collocations between the two words, *the Coronavirus* and *the virus* (indicated in the table 3 below). The highly ranked collocations between the two corpora demonstrate the commonality shared between these two words but their noticeably different orderings are much more

revealing.

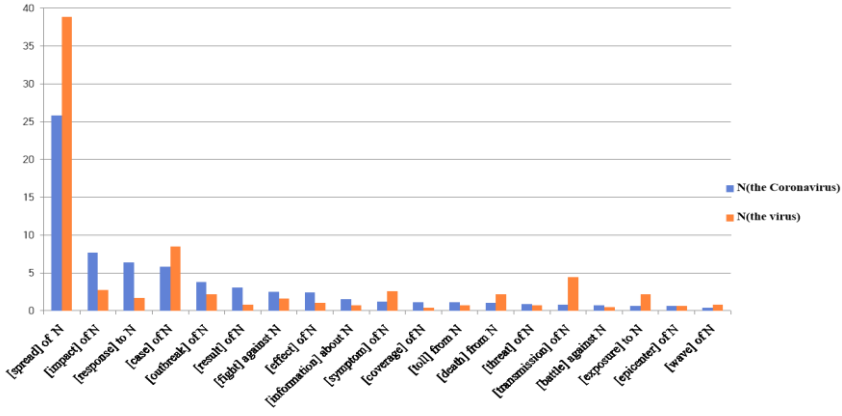
Table 3. List of all collocations (sorted by frequency)

	the Coronavirus (Coronavirus corpus)		*the virus* (NOW corpus)	
	Types	Tokens	Types	Tokens
1	[spread] of the Coronavirus	28083	[spread] of the virus	39434
2	[impact] of the Coronavirus	8335	[case] of the virus	8650
3	[response] to the Coronavirus	7014	[transmission] of the virus	4477
4	[case] of the Coronavirus	6304	[impact] of the virus	2779
5	[outbreak] of the Coronavirus	4169	[symptom] of the virus	2681
6	[result] of the Coronavirus	3367	[exposure] to the virus	2314
7	[fight] against the Coronavirus	2728	[death] from the virus	2299
8	[effect] of the Coronavirus	2716	[outbreak] of the virus	2262
9	[fallout] from/of the Coronavirus	1936	[response] to the virus	1746
10	[handling] of the Coronavirus	1878	[fight] against the virus	1649
11	[concern] about/over the Coronavirus	1814	[strain] of the virus	1455
12	[information] about the Coronavirus	1634	[effect] of the virus	1021
13	[symptom] of the Coronavirus	1339	[result] of the virus	908
14	[coverage] of the Coronavirus	1266	[risk] of the virus	862
15	[toll] from the Coronavirus	1200	[wave] of the virus	812
16	[death] from the Coronavirus	1144	[toll] from the virus	805
17	[update] on the Coronavirus	1122	[information] about the virus	758
18	[threat] of the Coronavirus	1064	[threat] of the virus	730
19	[transmission] of the Coronavirus	958	[source] of the virus	716
20	[battle] against the Coronavirus	817	[presence] of the virus	675
21	[exposure] to the Coronavirus	700	[epicenter] of the virus	661
22	[epicenter] of the Coronavirus	696	[carrier] of the virus	635
23	[story] about the Coronavirus	691	[test] for the virus	566
24	[experience] with the Coronavirus	627	[origin] of the virus	557
25	[start] of the Coronavirus	608	[battle] against the virus	515
26	[wave] of the Coronavirus	485	[nature] of the virus	515
27	[vaccine] for the Coronavirus	443	[sign] of the virus	509
28	[onset] of the Coronavirus	427	[coverage] of the virus	500
29	[fear] of the Coronavirus	425	[surface] of the virus	474
30	[recovery] from the Coronavirus	399	[resurgence] of the virus	461
Total		108644		101380

Also, it is interesting how [fallout] and [handling] as exclusive collocates of the keyword *the Coronavirus* rank high. The differences among these commonly shared collocations are explicitly laid out in the following chart 3.

It is shown from the chart 3 that *the Coronavirus* shares slightly closer affinity with the noun [impact], [response], [outbreak], [result], [fight], [effect] while *the virus* shares closer affinity with the noun [spread], [case], [symptom], [death], [transmission] and [exposure]. The results allow us to see a more finely grained difference even among these ostensibly common collocations. The list below further separates the collocations exclusive to each word and renders conclusive the differences already adumbrated in the foregoing sections.

Chart 3. Commonly identified collocations (words in yellow, sorted by ratio)



List 3. Exclusive collocations (words in green & red)

the Coronavirus	fallout from/of, handling of, concern about/over, update on, story about, experience with, start of, vaccine for, onset of, fear of, recovery from
the virus	strain of, risk of, source of, presence of, carrier of, test for, origin of, nature of, sign of, surface of, resurgence of

What this list 3 shows is that the nouns distantly collocating with *the Coronavirus* (e.g., [fallout], [concern], [update], [story], [experience], [recovery]) are the type of nouns that reflect direct ramifications of the Coronavirus outbreak instead of nouns that illustrate a certain property of the Coronavirus itself. The prepositions that these types of nouns select vary accordingly, which is in stark contrast to the uniformly selected preposition 'of' in noun collocations of the reference word *the virus*. This prevalence of the construction [N of *the virus*] implies that the occurrence of a specific noun (e.g., [strain], [carrier], [nature]) is only contingent on the reference word *the virus* in that the nouns filling in this slot are purposely employed to emphasize an aspect directly pertaining to the virus. In comparison, the occurrence of mutually unrelated nouns that nonetheless collocate strongly with the keyword

the Coronavirus shows how the focus is not laid on understanding the exact nature of the Coronavirus but rather on the diverse 'human endeavors' that only obliquely characterize such new virus. This in turn creates the impression that we are failing to go all-out grappling with the Coronavirus itself in being oblivious to its very nature in the first place.

3.2. Adjective Collocation

The immediately adjacent adjectives modifying the Coronavirus and the general virus were retrieved using the search string **'[ADJ] Coronavirus'** and '**[ADJ] virus'** in each corpus with the LIST search feature. There were a few complications with regard to selecting the tokens most relevant for the present study since a considerable amount of the adjective tokens were used to modify not only the Coronavirus or the virus but the entire compound noun such as those illustrated in the string [national] [[Coronavirus] command] or the [experimental] [[Coronavirus] vaccine]. This echoes the finding from the previous section which showed how freely *the Coronavirus* or *the virus* cooccur with other nouns. Since these types of adjectives clearly do not index an inherent property of the Coronavirus but modify the next occurring nouns, they were regarded irrelevant and abandoned. Some other adjectives displayed ambiguous scope in their modification as exemplified in the string [latest] [Coronavirus situation] or the [evolving] [Coronavirus outbreak]. Therefore, although the point still remains moot, a disclaimer is in order that I rather liberally included these types of tokens such as [latest] and [evolving] in the sorting process as I deemed them still tangentially relevant. Also, the determiner *the* in the search string has been deleted since it occupies the exact slot where immediately preceding adjectives of the two words would appear.

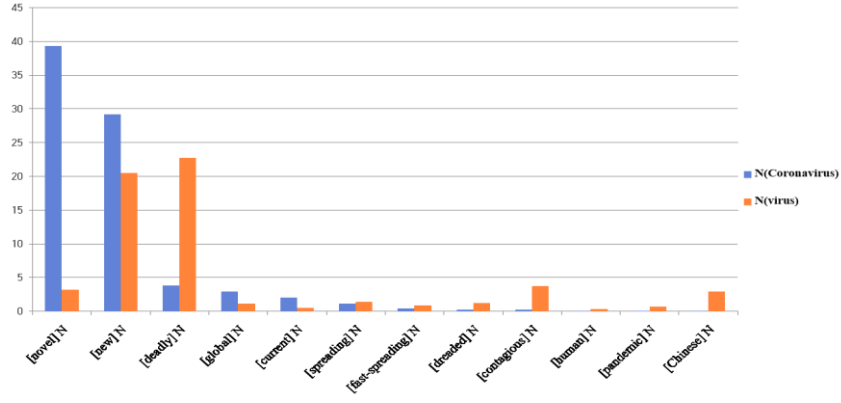
3.2.1. [ADJ] + *Coronavirus/virus*

Table 4. List of all collocations (sorted by frequency)

	Coronavirus (Coronavirus corpus)		*virus* (NOW corpus)	
	Types	**Tokens**	**Types**	**Tokens**
1	[novel] Coronavirus	68720	[deadly] virus	11457
2	[new] Coronavirus	51115	[new] virus	10383
3	[deadly] Coronavirus	6801	[contagious] virus	1891
4	[global] Coronavirus	5079	[novel] virus	1618
5	[ongoing] Coronavirus	3725	[respiratory] virus	1607
6	[current] Coronavirus	3607	[mosquito-borne] virus	1484
7	[spreading] Coronavirus	2056	[Chinese] virus	1475
8	[latest] Coronavirus	1803	[infectious] virus	886
9	[growing] Coronavirus	854	[flu-like] virus	826
10	[fast-spreading] Coronavirus	773	[syncytial] virus	820
11	[recent] Coronavirus	495	[spreading] virus	744
12	[known] Coronavirus	493	[dreaded] virus	650
13	[dreaded] Coronavirus	489	[particular] virus	624
14	[reported] Coronavirus	479	[live] virus	605
15	[active] Coronavirus	436	[global] virus	514
16	[contagious] Coronavirus	416	[lethal] virus	508
17	[raging] Coronavirus	406	[dangerous] virus	507
18	[severe] Coronavirus	329	[cold] virus	499
19	[human] Coronavirus	319	[fast-spreading] virus	455
20	[worldwide] Coronavirus	305	[sars-like] virus	441
21	[worst] Coronavirus	292	[terrible] virus	371
22	[pandemic] Coronavirus	281	[pandemic] virus	363
23	[worsening] Coronavirus	264	[airborne] virus	318
24	[Chinese] Coronavirus	251	[unknown] virus	313
25	[imported] Coronavirus	235	[mysterious] virus	283
26	[escalating] Coronavirus	209	[fatal] virus	273
27	[widespread] Coronavirus	187	[current] virus	268
28	[evolving] Coronavirus	184	[adeno-associated] virus	261
29	[ravaging] Coronavirus	169	[human] virus	185
30	[asymptomatic] Coronavirus	157	[nasty] virus	177
Total		174718		50423

In general, there were 12 collocation types found to be commonly occurring between the two words, with the remaining 18 collocations being exclusive to each word (indicated in table 4 above). If we look at the highly ranked adjectives modifying each word, there already exists a significant ordering difference even among these common collocations. Also noteworthy is the emergence of exclusive collocations ranked high above, which altogether adumbrates the contrasting characterization between the Coronavirus and the general virus. First, the contrasting collocation valences among the commonly identified collocations are displayed in the chart below.

Chart 4. Commonly identified collocations (words in yellow, sorted by ratio)



It is revealed from the chart 4 above that *Coronavirus* collocates more frequently with adjectives such as [novel] and [new] while *virus* predominantly collocates with the adjectives [deadly], [contagious] and [Chinese] albeit with strikingly different ratios. Even though all these adjectives are employed for the characterization of the two viruses, the 'lethality' of the general virus is heavily accentuated whereas the 'unfamiliarity' of the Coronavirus is brought to the fore in its description.

List 4. Exclusive collocations (words in green & red)

Coronavirus	ongoing, latest, growing, recent, known, reported, active, raging, severe, worldwide, worst, worsening, imported, escalating, widespread, evolving, ravaging, asymptomatic
virus	respiratory, mosquito-borne, infectious, flu-like, syncytial, particular, live, lethal, dangerous, cold, sars-like, terrible, airborne, unknown, mysterious, fatal, adeno-associated, nasty

From the list 4 above, it appears that the adjectives modifying *Coronavirus* are ambiguous in their scope in that it remains unclear whether they are characterizing the Coronavirus or the noun immediately following the Coronavirus. This equivocal usage of the

adjectives relates to the already attested extensive noun collocations of the Coronavirus. On this note, the adjectives such as [ongoing], [latest], [growing], [recent], [reported], [worldwide], [escalating], [widespread] short-circuit the Coronavirus to modify next following nouns such as [outbreak], [epidemic], or [crisis] (not cited directly in this section for space reason). This ambiguous characterization of the Coronavirus could also engender the misrepresentation of the Coronavirus in the media since it creates the illusion that the Coronavirus is the one being 'recent', 'growing', 'imported', or 'evolving' when it actually refers to the 'recent outbreak', 'growing crisis', 'imported patients' or 'evolving situation'. Only few adjectives such as [raging] and [ravaging] seem to characterize the Coronavirus in the absence of any following noun strings. This lack of adjectives 'exclusively' modifying the Coronavirus is partly attributable to the dual nature of the Coronavirus itself in that its lethality is not the source of trouble despite its far-reaching effects. This contradiction may be a factor that adds to its 'elusiveness', which the media may have been compelled to circumvent by adopting fuzzy adjectives. In contrast, the adjectives modifying the *virus* are relatively clear cut in that they capture the very nature of the virus in technical terms (e.g., [respiratory], [mosquito-borne], [flu-like]) or its widely recognized lethality (e.g. [infectious], [dangerous], [lethal]).

3.3. Verb Collocation

Initially, the verb collocations were searched using the COLLOCATION search feature (window left & right 1) of each corpus to reduce intervening adjective strings. In addition, as already amply mentioned in the foregoing sections, some nouns immediately followed both *the Coronavirus* and *the virus* so some verb tokens appeared to be idiomatically linked to those following nouns instead (i.e. [weather] [[the Coronavirus] storm]), which I still took to be relevant tokens considering the fact that these verbs still contribute to the

characterization of the Coronavirus or the virus. After the preliminary results, the LIST search feature with the search string '**[VERB]** ***the Coronavirus/virus*** **[(NOUN)]**' was used to retrieve verb types that took each word as their thematic objects. In the subsequent delimitation process were abandoned verb types such as [treat] since they were split in meaning such as in 'treat the Coronavirus as [ ]' and 'cure'. Some verbs tokens such as [keep], [bring] and [take] that are used in varied phrasal constructions proved extremely difficult to tease apart all their usages and were therefore abandoned. Also, their generally frequent occurrences throughout the entire corpora would make it hard to ascribe the prevalence of these verb tokens only in relation to the key words, the Coronavirus or the virus. Also, some verbs tokens such as [say], [believe] and [hope] that introduce a subordinate clause (with a deleted compromiser 'that') were abandoned as they do not directly take the Coronavirus or the virus as their objects. Causative verbs such as [make] or [let] were also excluded because of their subsidiary role in introducing the main verb. This meticulous delimitation process rendered only content verbs with clearly discernible meaning contribution. The next process was retrieving verb tokens using the search string '***the Coronavirus/virus*** **[VERB] ([NOUN])**', which placed the two words in the subject positions this time. Because the key word *the Coronavirus* or *the virus* could also have been embedded within a prepositional phrase to form a bigger noun phrase (i.e. the *cases* of the Coronavirus *rose*), the verbs only obliquely associated with these two words (i.e. *rose, reached, spiked*) were accordingly abandoned. Finally, the optionality of the following noun would make relevant the intransitive verb/transitive verb distinction, which could prove equally revealing in the characterization of the Coronavirus and general virus.

3.3.1. [VERB] + *the Coronavirus/virus* + ([NOUN])

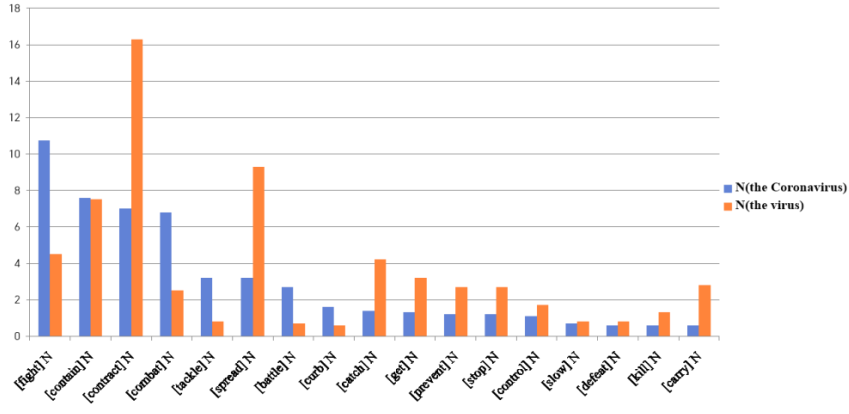
Table 5. List of all collocations (sorted by frequency)

	the Coronavirus (Coronavirus corpus)		*the virus* (NOW corpus)	
	Types	Tokens	Types	Tokens
1	[fight] the Coronavirus	4642	[contract] the virus	14358
2	[contain] the Coronavirus	3317	[spread] the virus	8178
3	[contract] the Coronavirus	3063	[contain] the virus	6622
4	[combat] the Coronavirus	2950	[fight] the virus	3964
5	[tackle] the Coronavirus	1399	[catch] the virus	3756
6	[spread] the Coronavirus	1389	[carry] the virus	3074
7	[battle] the Coronavirus	1204	[get] the virus	2859
8	[declare] the Coronavirus	772	[transmit] the virus	2709
9	[address] the Coronavirus	717	[prevent] the virus	2408
10	[curb] the Coronavirus	711	[stop] the virus	2378
11	[handle] the Coronavirus	684	[combat] the virus	2278
12	[catch] the Coronavirus	643	[control] the virus	1497
13	[call] the Coronavirus	602	[kill] the virus	1181
14	[use] the Coronavirus	586	[pass] the virus	1114
15	[get] the Coronavirus	583	[detect] the virus	839
16	[prevent] the Coronavirus	551	[slow] the virus	762
17	[stop] the Coronavirus	520	[tackle] the virus	726
18	[survive] the Coronavirus	514	[defeat] the virus	707
19	[control] the Coronavirus	505	[battle] the virus	699
20	[discuss] the Coronavirus	495	[eliminate] the virus	686
21	[monitor] the Coronavirus	373	[suppress] the virus	640
22	[manage] the Coronavirus	362	[curb] the virus	592
23	[cover] the Coronavirus	353	[eradicate] the virus	548
24	[slow] the Coronavirus	321	[beat] the virus	525
25	[weather] the Coronavirus	319	[acquire] the virus	416
26	[defeat] the Coronavirus	297	[track] the virus	413
27	[kill] the Coronavirus	285	[clear] the virus	377
28	[carry] the Coronavirus	284	[understand] the virus	370
29	[stem] the Coronavirus	277	[find] the virus	345
30	[face] the Coronavirus	251	[identify] the virus	343
Total		**43219**		**87888**

As indicated in the table 5 above, there were 17 commonly shared collocations (in yellow) with the remaining 13 collocations exclusive to each word. The highest ranks are mostly occupied by these commonly shared collocations but their relative ordering between these two corpora hardly coincide. This difference is again captured in the chart 5 below.

First, note how the most frequent verb collocation of *the virus* is [contract] while the most frequent verb collocation of its counterpart *the Coronavirus* is [fight]. A closer inspection shows us that *the Coronavirus* shares a closer semantic affinity with militant verbs such as [fight], [combat], [tackle], [battle] whereas *the virus* combines more robustly with a wider range of verbs such as [contract], [spread], [catch], [get], [prevent], [stop], [control], [kill], and [carry] that emphasize the transmission process of the virus and palliative measures

Chart 5. Commonly identified collocations (words in yellow, sorted by ratio)



in response. This difference becomes more definite if we look at the set of verbs that exclusively precede each word.

List 5. Exclusive collocations (words in green & red)

the Coronavirus	declare, address, handle, call, use, survive, discuss, monitor, manage, cover, weather, stem, face
the virus	transmit, pass, detect, eliminate, suppress, eradicate, beat, acquire, track, clear, understand, find, identify

The list 5 above shows the striking discrepancy between the verbs employed to characterize *the Coronavirus* and *the virus*. The verbs employed to describe *the virus* (e.g., [detect/ eliminate/ suppress/ eradicate/ beat/ track/ clear/ understand/ find/ identify]) attest to how we recognize the virus as a familiar entity that we can easily bring under control, which is a reflection of our past success in subjugating the virus and our unwavering confidence. The fact that the word *virus* mostly appeared as the verb's sole and direct object without any nouns appended to it syntactically bears out this implication.

In contrast, many verbs employed to describe *the Coronavirus* were

shown to take not only the Coronavirus but also the noun following the Coronavirus (or the entire compound noun) as their objects. Some example verbs that treated the Coronavirus as their sole object include [address], [handle], [survive], [discuss], [monitor], [manage], and [face]. These types of verbs reflect our heightened awareness, reflections, and implicit resilience towards the Coronavirus instead of a more head-on confrontation with the virus itself. Therefore, they demonstrate how we were pushed to take a more guarded stance in dealing with the Coronavirus due to its unprecedented pervasiveness. Nevertheless, still underlying the prevalence of these types of verbs is our perception of the Coronavirus as an unwelcome 'disturbance' that we can at least contain, which is an idealized faith even when the stark reality (i.e. absence of any resolute remedy) renders it futile. That is because the expressions such as [discuss/ handle/ monitor/ manage *the Coronavirus*] are entirely dismissing the agency of the Coronavirus, which in fact operates independently of the human will. It evokes the age-old endeavor to claim primacy over general virus and our reluctance to succumb to it, which falls squarely in line with the traditional view that casts natural virus as an object of conquest. In conclusion, what the verbs describing the Coronavirus hint at is our renewed profession of this long-standing belief and our unfailing hope that we still hold firm as 'maintainers' who can restore back the order.

3.3.2. *the Coronavirus/virus* + [VERB] + ([NOUN])

The result rendered from this search string showed that there are 14 commonly identified collocations (in yellow) with the remaining 16 collocations exclusive to each word (shown in table 6 below). However, what is most noteworthy is that the highly ranked collocations were largely in congruence with respect to their relative orderings. This also alludes to the fact that when *the virus* and *the Coronavirus* are postulated as agents, the actions they take are primarily formulated on

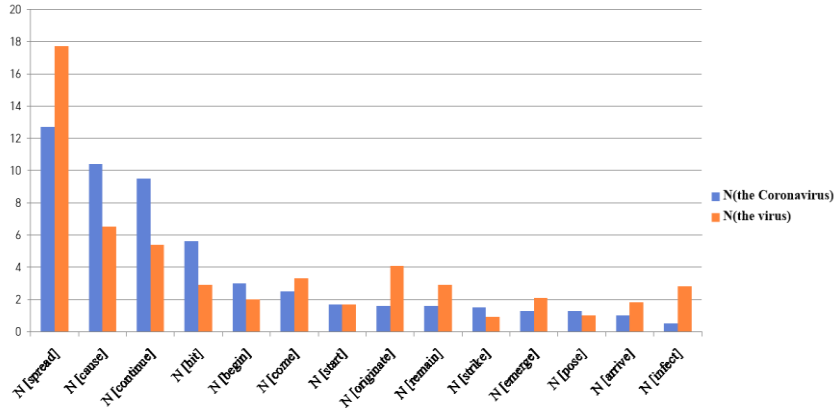
the basis of their communal attribute as a 'virus'. Another interesting fact is that intransitive verbs such as [spread], [continue], [begin], [come], [start], [originate], and [remain] are ranked high in the characterization of both viruses, which indicates that viral reproduction is the first and foremost feature in their characterizations.

Table 6. List of all collocations (sorted by frequency)

	the Coronavirus (Coronavirus corpus)		*the virus* (NOW corpus)	
	Types	Tokens	Types	Tokens
1	the Coronavirus [spread]	2065	the virus [spread]	5482
2	the Coronavirus [cause]	1702	the virus [cause]	2039
3	the Coronavirus [continue]	1544	the virus [continue]	1688
4	the Coronavirus [hit]	920	the virus [originate]	1286
5	the Coronavirus [begin]	496	the virus [come]	1049
6	the Coronavirus [come]	406	the virus [hit]	920
7	the Coronavirus [start]	283	the virus [remain]	899
8	the Coronavirus [originate]	268	the virus [emerge]	670
9	the Coronavirus [remain]	267	the virus [begin]	634
10	the Coronavirus [strike]	246	the virus [start]	550
11	the Coronavirus [shut]	226	the virus [enter]	462
12	the Coronavirus [force]	224	the virus [infect]	438
13	the Coronavirus [emerge]	217	the virus [pose]	323
14	the Coronavirus [pose]	210	the virus [strike]	287
15	the Coronavirus [prompt]	201	the virus [arrive]	285
16	the Coronavirus [arrive]	172	the virus [kill]	275
17	the Coronavirus [sweep]	150	the virus [rise]	232
18	the Coronavirus [threaten]	122	the virus [attack]	232
19	the Coronavirus [lead]	121	the virus [stay]	214
20	the Coronavirus [put]	113	the virus [survive]	192
21	the Coronavirus [break]	99	the virus [mutate]	172
22	the Coronavirus [impact]	94	the virus [replicate]	157
23	the Coronavirus [upend]	93	the virus [grow]	156
24	the Coronavirus [change]	89	the virus [move]	155
25	the Coronavirus [infect]	82	the virus [occur]	154
26	the Coronavirus [ravage]	81	the virus [run]	153
27	the Coronavirus [disrupt]	67	the virus [persist]	143
28	the Coronavirus [wreak]	60	the virus [claim]	137
29	the Coronavirus [require]	44	the virus [follow]	136
30	the Coronavirus [represent]	37	the virus [circulate]	117
Total		**16223**		**30943**

Despite this congruence, there still exist subtle differences in their ordering which are again plotted in the chart 6 below for further discussion.

Chart 6. Commonly identified collocations (words in yellow, sorted by ratio)



The commonly identified verbs again display different degrees of association with each word, *the Coronavirus* and *the virus*. It is shown that the keyword *the Coronavirus* has verb types such as [cause], [continue] and [hit] in its closer proximity while *the virus* has verb types such as [spread], [come], [originate], [remain], [emerge], and [infect] in its closer proximity. This different collocation pattern illustrates that in describing the Coronavirus, emphasis is more finely calibrated towards the concrete 'actions' it implements, and accordingly the consequences it brings about. This difference becomes more pronounced if we take a look at the verb types that emerge as exclusive collocates of each word in the list 6 below.

List 6. Exclusive collocations (words in green & red)

the Coronavirus	shut, force, prompt, sweep, threaten, lead, put, break, impact, upend, change, ravage, disrupt, wreak, require, represent
the virus	enter, kill, rise, attack, stay, survive, mutate, replicate, grow, move, occur, run, persist, claim, follow, circulate

The verb types in the list 6 above conclusively and substantially highlight the different perspective that we adopt in approaching the Coronavirus in comparison to general virus. First, the verbs characterizing general virus are limited to the ones either denoting its autonomous actions (e.g., [stay/ survive/ mutate/ replicate/ grow/ move/ occur/ run/ persist/ circulate]) or its lethality in terms of bodily infiltration (e.g., [enter/ kill/ attack/ claim]), which confirms findings from a previous research that demonstrated how a more lethal SARS virus was conceptualized as a single unified KILLER (Wallis & Nerlich, 2005).

In contrast, the Coronavirus is granted a markedly upgraded agency throughout its characterization as exemplified in the verbs [shut], [force], [sweep], [threaten], [lead], [put], [break], and [change] with heightened 'personification'. The destructive power of the Coronavirus is aggrandized in terms of its insidious penetration into every sphere of our lives. Not only do they [ravage], [disrupt], or [wreak (havoc)] but also navigate the globe ([sweep]), usher us into a new era ([lead]), engender an effect ([prompt/ upend/ change]) or at the extreme end of the spectrum develop to encode its own symbolic value ([represent]). Therefore, its pervasiveness is embodied through the multifarious activities it carries out, which derives from how the media animates the Coronavirus as a juggernaut that drives social changes. The Coronavirus as this catalyst of change thus surfaces as a prominent figure in the overarching Coronavirus 'narrative', with the humans inevitably relegated to being 'victimized' figures. In light of the findings from the previous section in which the Coronavirus was cast as a force we can triumph over, the dichotomy between the unruly Coronavirus and the debilitated human seems ostensibly self-contradictory, which is in and of itself a sad reflection of the current state of affairs and more importantly, our resolve to soldier on regardless.

4. Conclusion

In this paper, I have organized my study into three sections that would allow me to scrutinize the co-occurrence patterns of the main keyword *'the Coronavirus'* and the reference word *'the virus'* in terms of discrete part of speech type collocations. In the first section that dealt with the noun collocations of these two words, it is shown that there is a distinct tendency of the noun Coronavirus being used as an adjective modifying its immediately following noun, all of which coalesce into a newly coined compound noun. This emergence of new compound nouns revealed that the crucial component *'the Coronavirus'* is marginalized in that its semantic weighting on the meaning of a new compound noun is only minimal. On account of its flexible combination with nouns from diverse domains, it appears to pose less semantic imposition than its counterpart *'the virus'* which only collocates with a restricted set of nouns. Because its inherent property as a virus is veiled, it becomes more transparent in relation to its neighboring elements. In this respect, the Coronavirus is even more extraordinary since it takes on a referential function that indicates 'causality' between the two nouns. The results from examining the adjective collocations in the next section converge on this finding in that they show how the collocating nouns instead of the key Coronavirus noun receive prominence from the adjectives. The absence of any definite adjectives that underscore the features of the Coronavirus in its own right is made more salient in comparison to the adjectives that consistently accentuate the lethality of general virus. In contrast to this feeble characterization of the Coronavirus in noun and adjective collocations, the verb collocations present themselves as interesting loci where the characteristics of the Coronavirus are recognizably substantiated. Our perception towards and treatment of the Coronavirus are pulled into view, namely our conflicting attitudes towards the Coronavirus in that it is portrayed as a force to be reckoned with while at the same time, a fundamental

disruption that needs to be brought under control. Our struggle to address the challenges and normalize our lives are reflected in the way we typecast ourselves as helpless victims and the Coronavirus as relentless invaders, which is a testament to our outright resistance to the unknown. This dichotomous distinction has already been well documented in the conceptual metaphor domain with Sontag (1989) illustrating that the metaphor 'disease is an invader' is habitually invoked in people's conceptualization, especially in envisaging the virus as the alien 'other' (1989: 99). His claims were empirically bolstered by some following studies that covered the SARS epidemic (Chiang & Duann 2007; Chung, 2011). The infiltration of this foreign substance represents not only a physical invasion but also an encroachment on human existence. This figurative view we hold towards the Coronavirus is in this respect a corollary of our encounter with an unfamiliar menace, and our reflexive resort to metaphorization as a means of conceptualizing such a menace.

References

Ashrafi-rizi, H., & Kazempour, Z. (2020). Information typology in coronavirus (COVID-19) crisis; a commentary. *Archives of Academic Emergency Medicine*, *8*, 1-9.

Chiang, W.Y., & Duann, R.F. (2007). Conceptual metaphors for SARS: 'War' between whom?, *Discourse and Society, 18*(5), 579-602.

Chung, S.F. (2011). A corpus-based study of SARS in English news reporting in Malaysia and in the United Kingdom. *International Review of Pragmatics, 3*, 270-293.

Kouzy, R., Jaoude, J. A., Kraitem, A., Alam, M. B. E., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E., Baddour, K. (2020). Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus, 3*, 12.

Krause, N. M., Freiling, I., Beets, B., Brossard, D. (2020). Fact-Checking as risk communication: The multi-layered risk of misinformation in times of COVID-19. *Journal of Risk Research,*

1-8.
Mejova, Y., & Kalimeri, K. (2020). Advertisers jump on coronavirus bandwagon: Politics, news, and business, available at http://arxiv.org/abs/2003.00923.
Motta, M., Stecula, D., & Farhart, C. (2020). How right-leaning media coverage of COVID-19 facilitated the spread of misinformation in the early stages of the pandemic in the U.S.. *Canadian Journal of Political Science*, 1-8.
Patel, K., & Binjola, H. (2020). Fake news swamping interpersonal communication in the times of CoronaVirus, available at SSRN: https://ssrn.com/abstract=3600129.
Pulido, C. M., Villarejo-Carballido, B., Redondo-Sama, G., & Gómez, A. (2020). COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information. *International Sociology, 35*(4), 377–392.
Rahardi, K. (2020). Building critical awareness of CoronaVirus-related news: Cyber-pragmatic study of Covid-19 hoaxes on social media. *International Journal of Advanced Science and Technology, 29*(6), 5398-5409.
Sharma, K., Seo, S., Meng, C., Rambhatla, S., Dua, A., & Liu, Y. (2020). Coronavirus on social media: Analyzing misinformation in Twitter conversations, available at http://arxiv.org/abs/2003.12309.
Sontag, S. (1989). *Illness as Metaphor: AIDS and Its Metaphors.* New York: Picador/Farrar, Straus and Giroux.
Wallis, P., & Nerlich, D. (2005). Disease metaphors in new epidemics: The UK media framing of the 2003 SARS epidemic. *Social Science and Medicine*, *60*(11), 2629-2639.
Wen, J., Aston, J., Liu, X., & Ying, T. (2020). Effects of misleading media coverage on public health crisis: A case of the 2019 novel Coronavirus outbreak in China. *Anatolia: An International Journal of Tourism and Hospitality Research*, 1-6.

Yejin Kim
yjk8095@snu.ac.kr